

平成 28 年度
学士学位論文

注目物体抽出を用いた
部分領域に基づく類似画像検索

Visual Key Image Retrieval using
Object Extraction based on Saliency

1170176 領内 あゆみ

指導教員 吉田 真一

2017 年 2 月 28 日

高知工科大学 情報学群

要 旨

注目物体抽出を用いた 部分領域に基づく類似画像検索

領内 あゆみ

ユーザがクエリとなる画像なしに名称不明のあるオブジェクトを含む画像を探すことのできる画像検索システムは、日々画像の枚数が増加するインターネット上において重要である。それに対して、ユーザに検索の手掛かりとしてビジュアルキーという特徴的な複数の画像を提示する画像検索がある。本研究では、Cheng らの顕著物体検出をこの画像検索の前処理に導入する。この手法により、切り取られた顕著物体領域画像から部分画像を生成し、それらの画像特徴量を取得、ビジュアルキーを選定することで、セマンティックギャップを低減し、検索性能の向上をねらう。評価は、インターネット画像データベースである ImageNet の画像を用いて、被験者による検索操作で、検索結果。評価実験の結果、注目物体を抽出する手法が、抽出しない手法に対して、適合率が 21.3%か 23.5%、再現率が 11.4%から 16.2%、F 値は 14.8%から 19.2%に向上することを示す。

キーワード 類似画像検索, 注目物体抽出, ビジュアルキー型画像検索

Abstract

Visual Key Image Retrieval using Object Extraction based on Saliency

RYONAI Ayumi

Image retrieval system, which enables users search images of unlabeled objects without having any similar images is important while a number of images on the Internet increases every day . Visual key image retrieval (VKIR) is a system that displays several characteristic images called visual keys as search queries for users. Thus, users can retrieve images without having similar images. We propose VKIR using salient object detection as a preprocessing. This method produces cropped salient object region images. These segmented images are used to extract features for visual key selection. We evaluate the proposed method by a subjective experiment using ImageNet dataset. Our experimental results show that our method improves the accuracy from 21.3% to 23.5%, the recall from 11.4% to 16.2% and the F-measure from 14.8% to 19.2%. The results demonstrate that our system improves the performance of VKIR.

key words content based image retrieval, visual key image retrieval, salient object detection

目次

第 1 章	序論	1
第 2 章	関連研究	3
2.1	ビジュアルキーを使った類似画像検索	4
2.2	クラスタリング手法	4
2.2.1	Ward 法	5
2.3	画像特徴量	6
2.3.1	$L^*a^*b^*$ 色空間	8
第 3 章	注目物体抽出を用いた部分画像検索の提案	9
3.1	注目物体 (Salient Object)	9
3.2	大域コントラストに基づいた注目物体抽出	10
3.2.1	注目物体領域の特定	11
3.2.2	注目物体抽出	13
3.2.3	生成された部分画像の比較	14
3.3	注目物体の抽出に成功・失敗した例	14
3.4	提案システム	15
第 4 章	評価実験	18
4.1	画像検索実験	18
4.1.1	実験手順・条件	18
4.1.2	実験結果	19
4.1.3	考察	20
第 5 章	結論	24

目次

謝辭 25

参考文献 26

目次

2.1	ビジュアルキーが抽出されるまでの過程	5
2.2	Ward 法のアルゴリズム	7
3.1	注目物体領域を抽出するまでの過程 , (a) 入力画像 , (b)Felzenszwalb の セグメンテーション手法により分割された領域ごとに塗り分けられた画 像 , (c) 領域コントラストから求めた顕著性を表したグレースケール画像 , (d)SaliencyCut を使い注目物体と背景に二値化した画像	11
3.2	注目物体から部分画像を生成するための前処理 , (a) 入力画像 , (b) マスク画 像 (二値画像) , (c) 抽出された注目物体領域 , (d) (c) に合わせてトリムし た入力画像	13
3.3	注目物体抽出ありのときの部分画像	14
3.4	注目物体抽出なしのときの部分画像	15
3.5	注目物体を抽出できた例	16
3.6	注目物体を抽出できなかった例	17
4.1	各カテゴリの画像の一例 および 実験におけるテスト画像	19
4.2	注目物体抽出ありのときのビジュアルキー	20
4.3	注目物体抽出なしのときのビジュアルキー	21
4.4	画像検索実験における平均適合率・平均再現率・F 値	22
4.5	それぞれの手法において各クラスに属する部分画像の枚数	23

第 1 章

序論

類似画像検索 (Image Retrieval) は、デジタル画像からなる大きなデータベースから、画像を探す技術である。類似画像検索は、ユーザがクエリとして示したテキストや画像を元に、一致する情報を持つ画像をデータベースから見つけ出し、ディスプレイに提示する。ソーシャルネットワークサービスやクラウドコンピューティングの発展により、インターネット上に存在する画像の数は年々増加している。文献 [1] によると、検索エンジンがインターネット上で検索可能な画像の情報量 (推計値) は 2009 年には 1,798TB にも上り、2003 年の 313TB と比較すると、6 年間で約 8 倍にも増大している。また、一人当たりの検索サービスの月間平均利用回数は、2002 年の 4.5 回から 2008 年には約 2 倍の 8.9 回に増加している [1]。このような状況下で、ユーザの要求に対して適切な検索結果を返すことがますます重要となる。

現在、主流の類似画像検索手法である Text-Based image retrieval は、アノテーションによって、あらかじめ各画像にメタデータとしてキーワードを付与しておき、データベースに格納しておく [2]。そして、ユーザから検索クエリが与えられたとき、画像検索エンジンがデータベース上でそのクエリに適合するキーワードを探す。そして、紐づけられた画像が関連度の順に結果として表示される。しかし、画像一枚一枚にキーワードを与えるアノテーションには、人手が必要で時間がかかるという欠点がある。また、紐付けるメタデータの適切さと全ての画像に必要なメタデータが与えられているかという網羅性によって、検索の精度が左右される。近年、ディープラーニングによって自動で画像へ説明文を与える手法 [3] も提案されているが、いまだに一般的には使われてはいない。それに対して、1992 年に加藤によって提案されたのが、Content-Based image retrieval (CBIR) である [4]。CBIR は

画像自身が有する情報を使って画像の類似度をはかる。CBIR では、ユーザからクエリとして与えられた画像の特徴量とデータベース上の画像の特徴量（色・テクスチャ・形など）の距離を測り、それが 0 に近いほど似ている画像であると判断される。このことから、人の主観によらずすべての画像を網羅的に比較することが可能となる。しかし、この手法でも、画像の特徴量と人が似ていると感じるものとの間に隔たりが生じることがあるという、セマンティックギャップの問題がある。その原因の一つとして考えられるのが、人間が視界の全てではなく、特定の箇所に注目しているという性質である。

そこで、本研究では、人間のその性質を模倣し、類似画像検索で画像の特徴量を取得する範囲を、画像中で最も目立つ物体である注目物体（salient object）がある領域に限定する。Salient Region Detection を適用して注目物体を抽出することにより、抽出しなかった場合とで画像検索の精度を比較する。

第 2 章では、類似画像検索の関連研究について記す。第 3 章では、画像特徴量を取得する前処理として適用する注目物体を抽出するための手法について記す。第 4 章では、実際に被験者に対して行った画像検索実験について記す。第 5 章で、本研究全体をまとめる。

第 2 章

関連研究

類似画像検索は、画像の類似性をはかる方式により、大きく分けて次の二種類に分類できる。

- Text-Based Image Retrieval (TBIR)
- Content-Based Image Retrieval (CBIR)

TBIR は、あらかじめデータベース上の全ての画像にメタデータとしてキーワードや文章を付与しておき、ユーザが単語やフレーズで画像検索を行った時に、関連度順に並べたサムネイル画像を結果として返すシステムである。現在の主要な画像検索サービスである Google Image Search などにも、これが用いられている。しかし、TBIR には、人手によるメタデータの付与に時間がかかったり、検索精度が各画像に割り当てられるメタデータの妥当性や網羅性に左右されたりするという欠点がある。一方、CBIR は、メタデータを付与せず、色やテクスチャ・形といった画像それ自身が持つ情報の類似性を基に画像を検索する。そのため、メタデータを与える場合の主観が排され、すべての画像を同一の観点で評価することができる。応用として、顔認識などに使われている。

また、CBIR には、画像特徴抽出、多次元データのインデクシング、検索システムの設計 (UI) という 3 つの基盤となる技術がある [5]。

本章ではそれらについて

1. 検索システムの設計：ビジュアルキー型画像検索
2. 多次元データのインデクシング：Ward 法によるクラスタリング

2.1 ビジュアルキーを使った類似画像検索

3. 画像特徴抽出： $L^*a^*b^*$ 色空間上のヒストグラム特徴

の順に述べる．

2.1 ビジュアルキーを使った類似画像検索

ビジュアルキー型画像検索は，2006年に Serata らによって提案されたものである [6]．この手法は，ユーザがクエリとなる画像を用意できないケースを想定したものである．この手法ではまず，データベースの全ての画像から，分割して部分画像を生成する．そして，それぞれの部分画像から画像特徴量を取得し，その特徴量をクラスタリングする．ここで，各クラスタの中心に最も近い特徴量を有する部分画像をビジュアルキーとする．次に，データベース中の各画像は，その部分画像が含まれるクラスタのビジュアルキーがインデックスとして付与される．実際の検索時には，ビジュアルキー（クラスタリングのクラスタと同数）がディスプレイに提示され，ユーザはそれらの中から，探している画像と部分的に一致すると考えるものを選択する．これにより，選択されたビジュアルキーをインデックスとして持つ画像が検索結果として表示される．

2.2 クラスタリング手法

クラスタリング（クラスター分析とも）とは，同じグループ内のデータは他のグループに含まれているデータよりも似通ったものになるように，一連のデータをグループ分けすることである．この分かれたグループをクラスタという．本研究では，ビジュアルキーの選定のために画像特徴量のクラスタリングを行う．

クラスタリング手法の例として，以下のものが挙げられる．

- 階層的クラスタリング：最短距離法，最長距離法，群平均法，Ward 法
- Centroid-based clustering：K-means 法
- Distribution-based clustering

2.2 クラスタリング手法

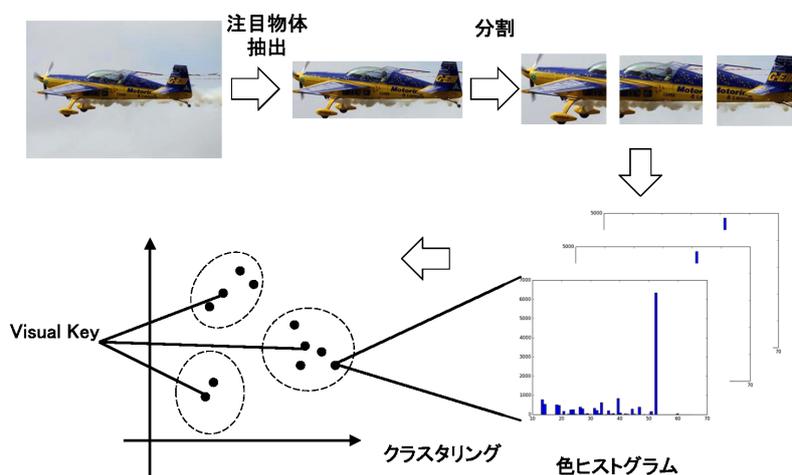


図 2.1 ビジュアルキーが抽出されるまでの過程

- Density-based clustering : DBSCAN (各クラスタの密度を同じにする)
- Affinity Propagation
- MeanShift
- Spectral Clustering
- Agglomerative Clustering
- Birch

本研究では階層的クラスタリングである Ward 法 [7] を用いる。

2.2.1 Ward 法

統計学において、Ward 法は階層的クラスタ分析の手法として用いられる。ベクトル空間において、「より近いオブジェクトは遠いオブジェクトより関係がある」ということに基づいたアルゴリズムである。応用例としては、データマイニングや機械学習、パターン認識など

2.3 画像特徴量

の統計的データ解析などがある。利点として、他の距離関数と比較して分類感度が高いことが挙げられる。K-means 法と比較すると、K-means 法は最初の点の取り方がランダムであるため、その取り方によってクラスが試行ごとに異なる。そのため、次元空間中におけるクラスターの分布が偏ってしまうことがある。それに対して Ward 法は、各要素間の距離を表す樹形図を始めに作っておき、それから決められたクラスター数に分割するため、異なる試行でも同一の結果を得ることができる。また、外れ値に対する頑健性も有している。式 2.1 は、クラスター C_m と C_n の距離 d_{mn} を表し、クラスター C_m と C_n を併合した際のクラスター内分散 $\text{Var}(C_m \cup C_n)$ が、クラスター C_m, C_n それぞれのクラスター内分散 $\text{Var}(C_m), \text{Var}(C_n)$ の和から、どの程度増えるかで定義する。

最小分散法

$$d_{mn} = \text{Var}(C_m \cup C_n) - (\text{Var}(C_m) + \text{Var}(C_n)) \quad (2.1)$$

Ward 法のアルゴリズムは以下の通りである。これを図で示したものが図 2.2 である。

1. 最初にクラスター中心を仮定せず、各オブジェクトをひとつのクラスターとみなす。
2. すべてのオブジェクト間の距離を計算し、その中で最も近い一組のオブジェクト対（オブジェクト 1 とオブジェクト 2）を見つける
3. その対をひとつのクラスター（オブジェクト 1・2）とみなす
4. またすべてのオブジェクト間の距離を計算する
5. 同様に最も近いクラスター対を併合・全てのクラスター間の距離の再計算を繰り返す
6. すべてのオブジェクトが一つのクラスターにまとまった時点でクラスターリングを終了する

2.3 画像特徴量

画像特徴量は、機械学習やパターン認識分野や画像処理分野において用いられる。画像の画素情報そのままでは処理するには情報量が多すぎたり、冗長であるとみなされるとき、別の形のより低次元の特徴ベクトルに変換される。与えられた画素情報の集合のうち、何を扱うかを決めることを特徴選択という。この特徴量は、目的とするタスクの遂行にあたって使

2.3 画像特徴量

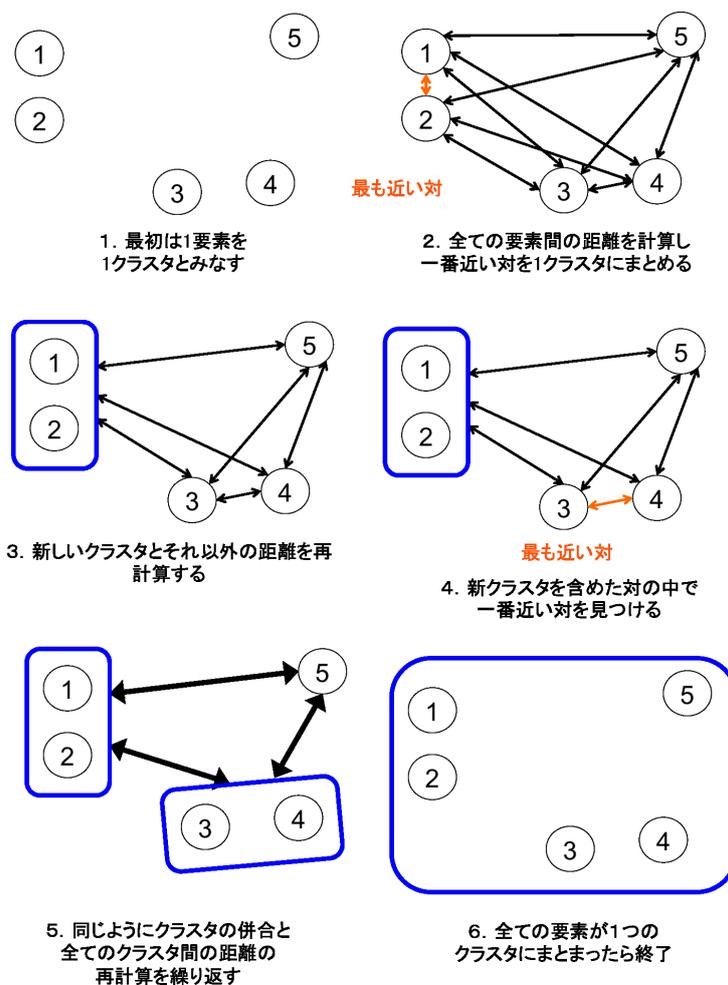


図 2.2 Ward 法のアルゴリズム

うことができるように、それぞれが元々の画像をよく表すものである必要がある。また、画像特徴量を利用するには更に次元削減を行うべきケースもある。

画像特徴量の例として、

- 形状特徴：Local Binary Pattern (LBP)[8]

2.3 画像特徴量

- Microstructure descriptor (MSD) : エッジの方向を基に微細な構造を抽出する [9]
- Global correlation descriptor (GCD) 色とテクスチャ[10]
- ORB
- Spatial position (空間的位置)
- Color-auto-correlogram (RGB)
- Gabor wavelet
- Wavelet transform

が挙げられる．本研究では， $L^*a^*b^*$ 色空間における色ヒストグラムを特徴量として用いる．

2.3.1 $L^*a^*b^*$ 色空間

$L^*a^*b^*$ 表色系は，1976 年に国際照明員会 (CIE) で定められた，色を定量的に表す体系のひとつである．色として定められている．特徴として，RGB 色モデルや CMYK 色モデルよりも，人間の視覚に近いものとして設計されている． $L^*a^*b^*$ のビットマップ画像が RGB や CMYK のビットマップと同じ精度を得るためには，一画素あたりより多くのデータを必要とする．また，明度を表す次元である L は 0~100， a^* および b^* はそれぞれ -128 ~ 127 の間の値をとる．一枚の画像中の最大輝度値を基に，全ての画素の相対輝度値が求められる．ある 2 色間は，

$$\sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad (2.2)$$

で求められる．

本研究では，この特徴量を取得する範囲を注目物体のある範囲に限定する．

第 3 章

注目物体抽出を用いた部分画像検索 の提案

本研究で類似画像検索に適用するのは，画像の中で最も目立っている物体である注目物体 (Salient Object) に対する物体検出で，2015 年に Cheng らによって提案された手法 [11] である．本章では注目物体という概念，Cheng らが提案した手法の手順について述べた後，注目物体の抽出の成功・失敗例について紹介し，最後に本研究で提案するシステムを説明する．

3.1 注目物体 (Salient Object)

Saliency は，従来は脳科学分野において用いられていた概念で，1985 年に Koch らが文献 [12] において，色や方位・動きといった複数の異なる視覚的特徴量を一つの信号に統合した分布図として，saliency map という概念を提案した．これは日本では顕著性マップとも呼ばれるものである．また，同論文の中で彼らは，“ある特定の場所についての saliency (顕著性) は，まずその地点が周囲と比べて色や方位・動き・奥行きなどにおいてどれだけ違っているかによって決まる”と述べている．顕著性とは，どれだけ目立っているか，人間がどれだけ注意を向けるか (注目するか) を表す値である．

注目物体に関する研究としては他に，Robinson らによって 1992 年に発表された視床枕内の神経で生じる信号と視覚的な saliency についての研究 [13] などがある．彼らの論文には，“彩り豊かな視覚を持って生きる上での最大の問題の一つは，適切に処理をしたり，注

3.2 大域コントラストに基づいた注目物体抽出

意を向けたりするために、その時々でどの物体が最も salient であるかを判断すること”とある。つまり、人間は生きていて無意識のうちに視界に映る物のうちの何かに注目しているということである。

この salient (saliency) の概念を工学の分野に導入したものが、視覚注意モデル分野についての研究のひとつである Salient Region Detection である。初期の研究としては、1998 年に Itti らにより提案された saliency based scene recognition が挙げられる [14]。彼らは、前述の saliency map のモデルをそのままシーン認識に適用し、輝度と色・方位 (orientation) を基に、場面中のすべての画素の顕著性 (saliency) をスカラー値で算出している。ここでは、輝度特徴量は 6 次元、色特徴量は 12 次元、方位の特徴量は 24 次元の計 42 次元の特徴量として計算される。現在では、Shen らの研究 [15] のように、region (領域) ではなく object (物体) レベルで顕著性に基づいたセグメンテーションができるようになっている。このことから、物体認識や画像編集・画像検索の前処理に応用されはじめている。

3.2 大域コントラストに基づいた注目物体抽出

これまでに注目領域の概念を画像検索に取り入れた研究として、2009 年の塩崎のものがある [16]。これは、デザイン学に基づいてエッジ抽出を用いた手法である。しかし、この手法では人間が背景として捉える部分も注目領域として抽出されることがあるという問題点がある。そこで本研究に取り入れたのは、2015 年に Cheng らによって提案された大域コントラストに基づいた注目物体の検出手法である [11]。この手法は注目物体のセグメンテーションの課題において、他の手法よりも高い 90% を超える適合率と F 値を結果として出すことに成功している。この手法の特徴は、

1. 大域的・空間的な色コントラストに基づく顕著性算出
2. 既存手法の GrabCut[17] を反復して適用する SaliencyCut

のふたつである。

3.2 大域コントラストに基づいた注目物体抽出

3.2.1 注目物体領域の特定

本小節では，入力された画像から白い注目物体領域と黒い背景領域に分かれた二値画像を生成する．

まず，グラフ理論に基づいた既存のセグメンテーション手法 [18] を使い，入力画像をいくつかの領域に分割する．入力画像である図 3.1(a) を分割すると，図 3.1(b) のようになる．これは，Felzenszwalb のウェブページ [19] で配布されているプログラムの実行によって生成されたもので，パラメータを $\sigma = 0.5$ ， $k = 500$ ， $min = 20$ としたものである．

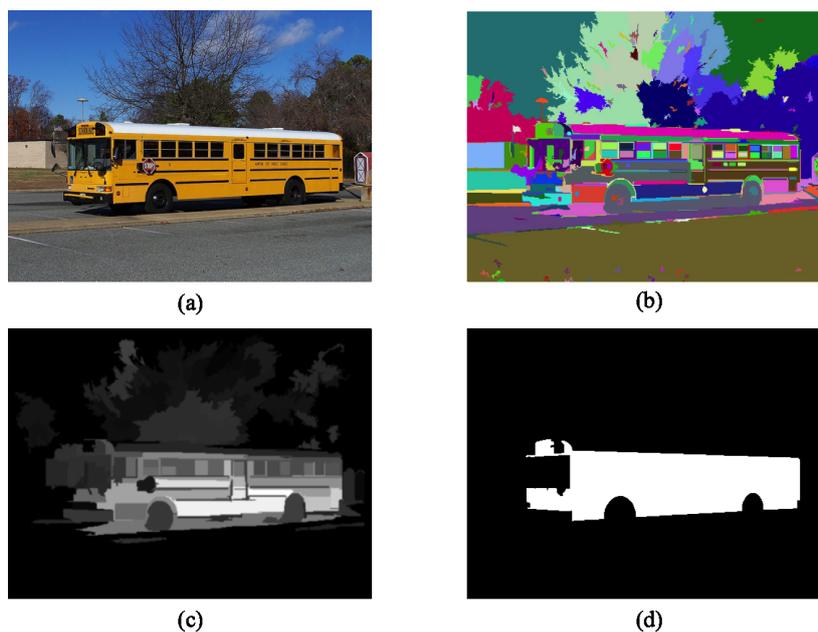


図 3.1 注目物体領域を抽出するまでの過程，(a) 入力画像，(b) Felzenszwalb のセグメンテーション手法により分割された領域ごとに塗り分けられた画像，(c) 領域コントラストから求めた顕著性を表したグレースケール画像，(d) SaliencyCut を使い注目物体と背景に二値化した画像

3.2 大域コントラストに基づいた注目物体抽出

次に，式 (3.1) によって，領域ごとの顕著性 (saliency) が求められる．この顕著性は，空間的重み付けをされた領域コントラストに基づいたものである． $w_s(r_k)$ は一つ目の空間重み付け項で，領域が画像の中心に近ければ大きい値をとり，遠ければ低い値をとる． $\exp(\frac{D_s(r_k, r_i)}{-\sigma_s^2})$ は二つ目の空間重み付け項であり，指数項の分子の $D_s(r_k, r_i)$ が領域のセントロイド間のユークリッド距離を表し，分母の $-\sigma_s^2$ が空間距離重み付けの強さを調整する役割をもつ．これにより，近い領域の影響を増やし，遠い領域の影響を減らすことができる． $w(r_i)$ は領域 r_k の重みで，領域 (r_k) に含まれる画素数により定義される． $D_r(r_k, r_i)$ は，二つの領域間の色差を表し，式 (3.2) によって求められる．したがって，重み付け項により，近い領域との色差が大きく画像の中心に近い領域ほど，より大きい顕著性の値を得ることになる． $f(c_{1,i})$ と $f(c_{2,j})$ はある色がその領域に占める割合である．

$$S(r_k) = w_s(r_k) \sum_{r_k \neq r_i} e^{\frac{D_s(r_k, r_i)}{-\sigma_s^2}} w(r_i) D_r(r_k, r_i) \quad (3.1)$$

$$D_r(r_1, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_{1,i}) f(c_{2,j}) D(c_{1,i}, c_{2,j}) \quad (3.2)$$

これらの式により，図 3.1(c) のように顕著性をグレースケールで表した画像が得られる．

最後に，GrabCut を改良した SaliencyCut を適用して画像を注目物体 (白) と背景 (黒) に二値化する．本来，GrabCut は手作業によって画像中で大まかに矩形で指定された物体を，その輪郭に沿って抽出するための手法である [17]．SaliencyCut は，Grabcut で必要であった人手によるアノテーションを不要にしたものであり，以下の 2 点を特徴とする．

1. アルゴリズムの初期化: 顕著性を表したグレースケール画像を閾値 T_b で二値化した画像を初期状態とする．
2. 反復フィッティングによるセグメンテーション: 最大 4 回反復して GrabCut を行う．反復することにより，より正確に注目領域を抽出する．このような反復過程は前述の顕著性算出アルゴリズムで得られるようなノイズの多い入力を扱うに当たって必要なものである．なぜなら，最初に設定した閾値から二値画像にした場合，画像によっては注目物体が背景としてセグメントされてしまうことがあるためである．

3.2 大域コントラストに基づいた注目物体抽出

これにより，注目物体領域が白，それ以外の部分が黒で示された図 3.1(d) のような二値画像が生成される．この図は github[20] で Cheng により配布されているプログラムによって生成されたものである．

3.2.2 注目物体抽出

本小節では，前小節で生成した二値画像を使い，入力画像から注目物体を抽出し，特徴量を取得するための部分画像を生成する．

まず，二値画像（図 3.2(b)）をマスクとして入力画像（図 3.2(a)）にマスク処理を施す．これにより，図 3.2(c) のように注目物体が抽出される．そして，画像検索に使えるように，

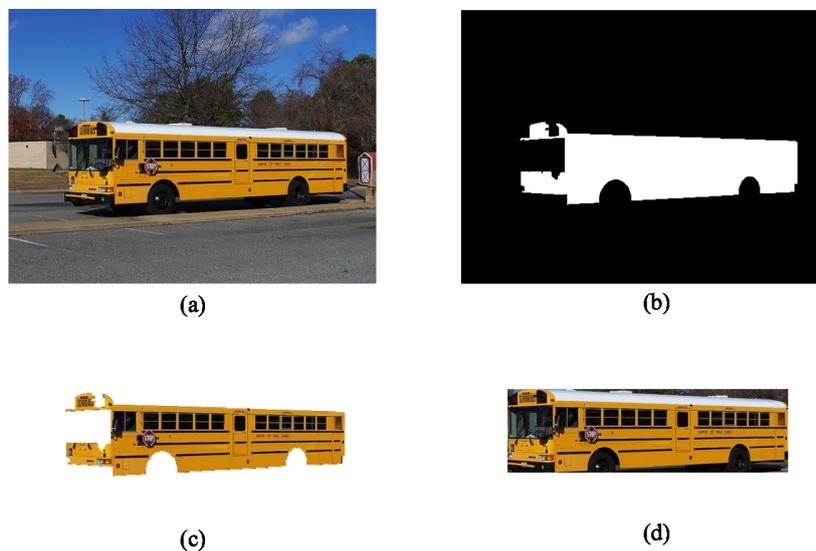


図 3.2 注目物体から部分画像を生成するための前処理，(a) 入力画像，(b) マスク画像（二値画像），(c) 抽出された注目物体領域，(d) (c) に合わせてトリミングした入力画像

抽出範囲がおさまる最小の矩形を，入力画像からトリミングして切り出す（図 3.2(d)）．こ

3.3 注目物体の抽出に成功・失敗した例

れにより，一枚の画像から注目物体領域を抽出することができる．

3.2.3 生成された部分画像の比較

前小節の手法により抽出された注目物体から，特徴量を取得するための部分画像を生成する．

図 3.3 が，注目物体を抽出してから分割することにより得られた部分画像である．図 3.4 が，注目物体を抽出せずに画像全体を分割することにより得られた部分画像である．入力画像一枚から得られる部分画像の枚数は，抽出ありで 3 枚，抽出なしで 12 枚と，注目物体を抽出したときの方が少なくなっている．また，抽出なしの場合では背景のみからなる部分画像ができています．



図 3.3 注目物体抽出ありのときの部分画像

3.3 注目物体の抽出に成功・失敗した例

Cheng らの注目領域抽出法はどのような画像に対しても有効なものではない．意図されていた注目物体でなく，背景の一部を抽出したり，全く何も抽出しないことがある．図 3.5 に注目物体を抽出できた例を，図 3.6 に注目物体を抽出できなかった例を示す．本研究で用

3.4 提案システム

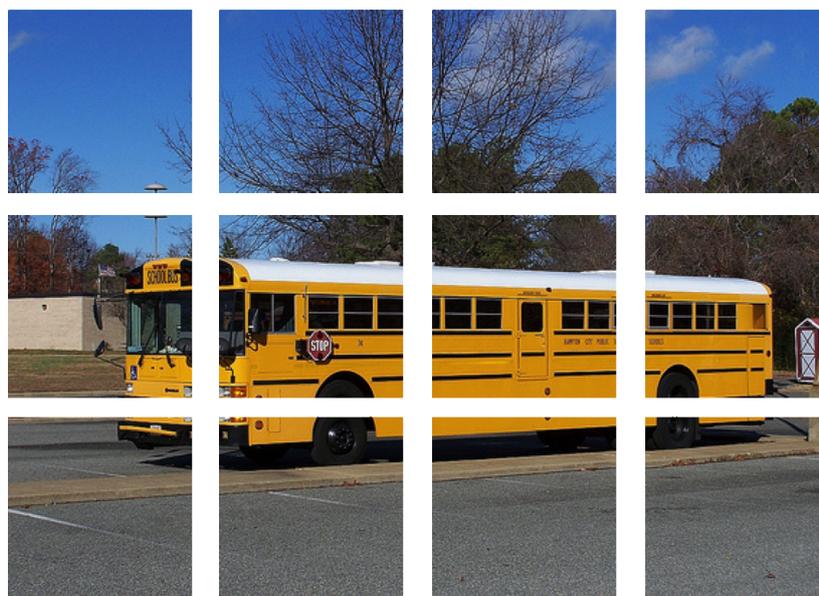


図 3.4 注目物体抽出なしのときの部分画像

いた 160 枚の画像データセット中，注目物体を抽出できたとみなすことのできた画像は 46 枚である．

3.4 提案システム

データベースのすべての画像（以後，元画像とする）から，注目物体抽出あり，および抽出なしそれぞれの手法で部分画像を生成する．生成する部分画像のサイズは 128×128 [pixel] とする．生成する際には部分画像同士の重複をゆるし，元画像 1 枚あたりに対して最小枚数となるようにする．生成されたすべての部分画像から，画像特徴を取得する．取得する画像特徴として， $L^*a^*b^*$ 色空間における色ヒストグラムを用いる．このヒストグラムは輝度値はそのままとし，補色次元の a^* ， b^* をそれぞれ 8×8 の 64 色に量子化したもので

3.4 提案システム

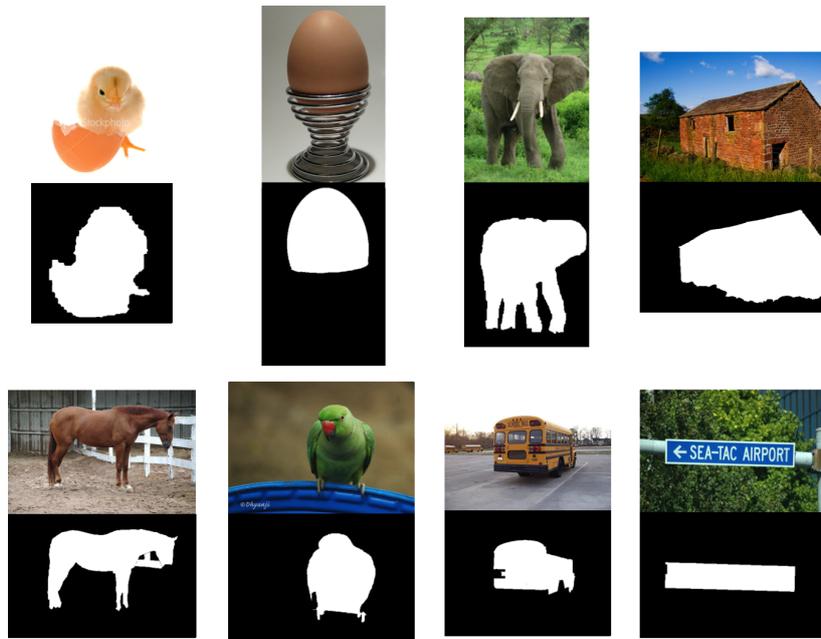


図 3.5 注目物体を抽出できた例

ある．そして，64 次元の L^*a^*b 色ヒストグラム空間において，Ward 法を使い部分画像をクラスタリングする．このときのクラスタ数は 20 とする．クラスタリングによって得られた各クラスタの重心（セントロイド）に最も近い部分画像をビジュアルキーとし，画像検索の際に被験者に提示する．

3.4 提案システム

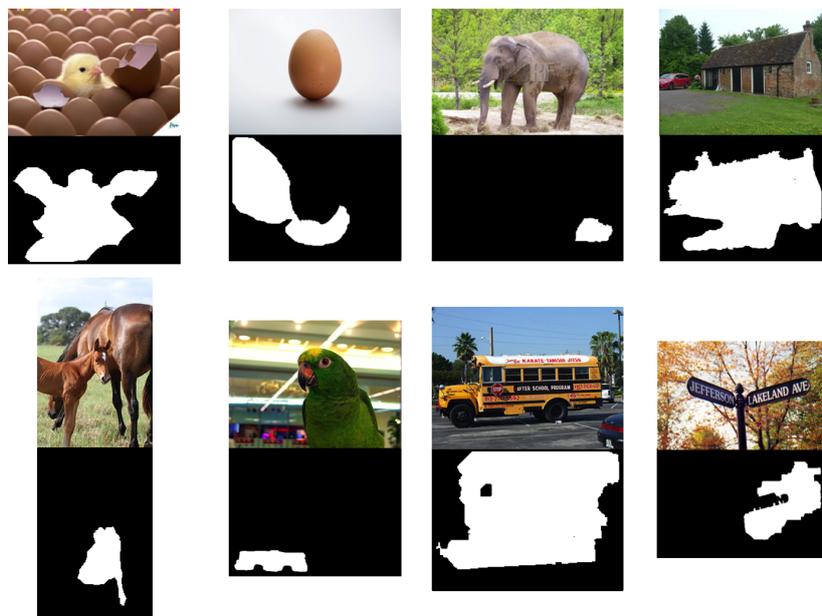


図 3.6 注目物体を抽出できなかった例

第 4 章

評価実験

本実験では、画像データセットに 3 章で述べた注目物体の領域抽出法を前処理として適用することで、部分領域に基づくビジュアルキー型類似画像検索における注目物体の抽出の有効性について調べる。評価指標として、適合率・再現率および F 値を算出する。適合率は、検索結果の画像のうち、テスト画像に類似する画像の占める割合であり、再現率は、データベース上でテスト画像に類似するもののうち、実際に検索結果として出てきた画像の割合である。また、F 値は適合率と再現率の調和平均として求める。

4.1 画像検索実験

ビジュアルキーを使った画像検索実験を実施し、画像の特徴量を、抽出された注目物体の領域から取得する場合と画像全体から取得する場合とで検索の精度を比較する。

4.1.1 実験手順・条件

実験用データとして、画像データベース ImageNet[21] の画像を用いる。Chick (n01792042), Egg (n07840804), Elephant (n02503517), Farm building (n03322570), Horse (n02374451), Parrot (n01816887), School bus (n02924116), Street sign (n06794110) の 8 カテゴリから、各 20 枚を用意する。(カテゴリ名の後ろのカッコ内はダウンロード時点での Wordnet ID を表す。) 各カテゴリの画像例を図 4.1 に示す。図 4.2 および図 4.3 に、被験者に提示したビジュアルキー(各手法 20 枚ずつ)を示す。図 4.2 は抽出した注目物体の部分画像をクラスタリングすることにより得られたビジュアルキーであ

4.1 画像検索実験

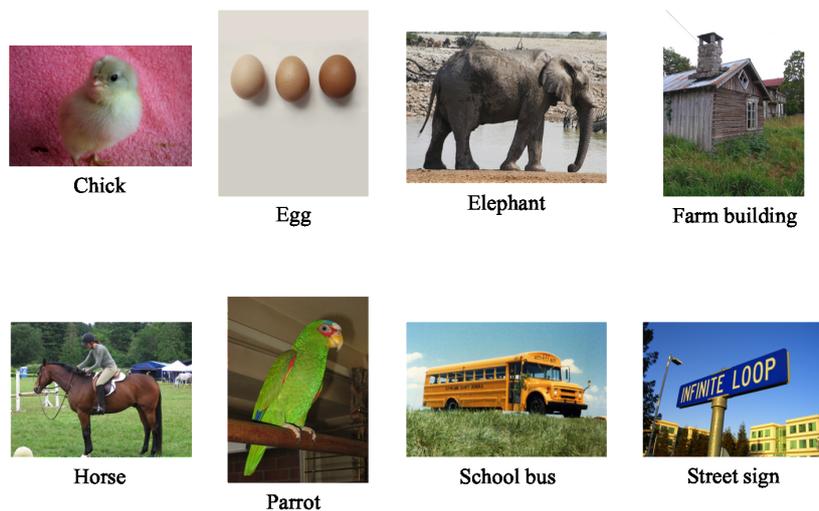


図 4.1 各カテゴリの画像の一例 および 実験におけるテスト画像

り，図 4.3 は画像全体の部分画像をクラスタリングすることにより得られたビジュアルキーである．6 名の被験者に対して実験を行う．提示されたテスト画像を検索するためのクエリとなるビジュアルキーを，各被験者が 20 枚のうちから 1 枚選ぶことを 1 試行とし，各手法 8 試行ずつ繰り返す．なお，テスト画像として 1 試行につき図 4.1 に示したうちの 1 枚が提示される．また，テスト画像が提示される順はすべての被験者において同一とする．

本実験では，色ヒストグラムを特徴量として類似性をはかることから，実験実施前に実験実施者が色が似ていると判断した画像群を類似しているものとして，適合画像集合として定めておく．この適合画像集合を基に，適合率および再現率を算出する．

4.1.2 実験結果

ビジュアルキーを使った画像検索実験において，全被験者の適合率の調和平均・全被験者の再現率の調和平均およびこれらふたつの値の調和平均である F 値を棒グラフにしたもの

4.1 画像検索実験

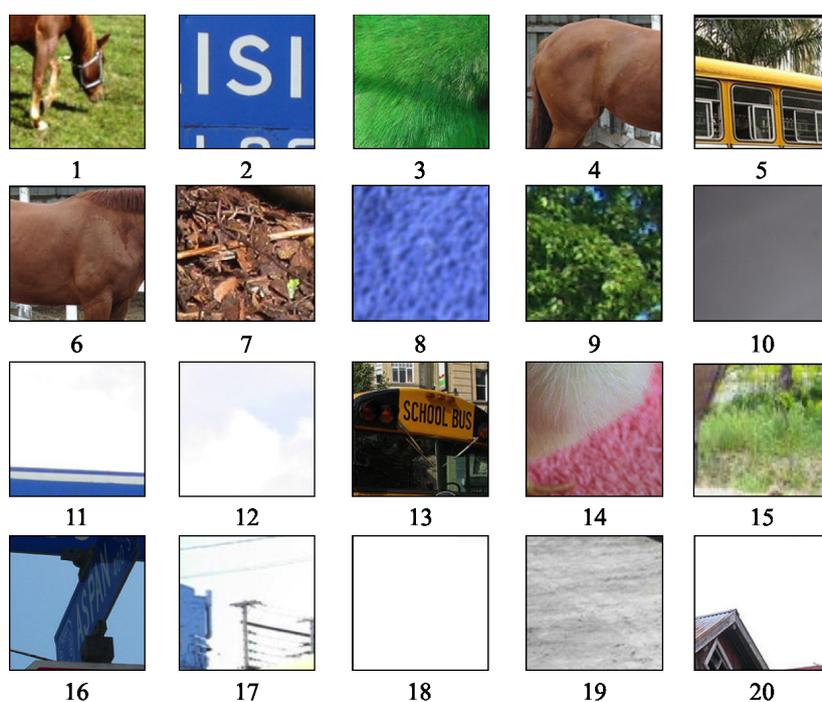


図 4.2 注目物体抽出ありのときのビジュアルキー

が図 4.4 である．縦軸が 8 回の試行における平均適合率および平均再現率を表している．抽出なしの平均適合率は 0.213，平均再現率は 0.114 であったのに対し，抽出ありの平均適合率は 0.235，平均再現率は 0.162 となり，平均適合率，平均再現率ともに抽出ありが抽出なしを上回る結果が得られた．このことから，注目物体抽出がビジュアルキー型画像検索において部分画像を生成する前処理として有用であると考えられる．

4.1.3 考察

まず，クラスタリングにおけるクラスタ数について考察する．本研究で行った実験で設定したクラスタ数は，これまでのビジュアルキー型類似画像検索についての研究を踏まえ 20

4.1 画像検索実験

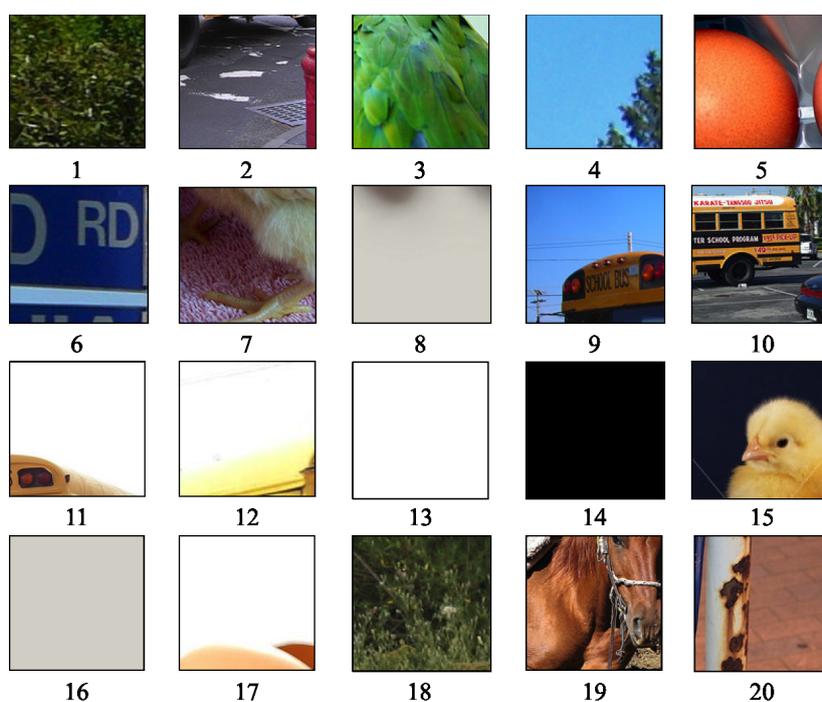


図 4.3 注目物体抽出なしのときのビジュアルキー

とした．しかし，図 4.5 は属する部分画像の枚数が極端に少ないクラスタの存在を示している．このことは，今回設定したクラスタの数が多すぎた可能性を示唆するものである．実際に，同系色が別のクラスタに分けられていることを確認することができた．これについては，人間が識別できる基本的な色カテゴリが赤・緑・青・黄・桃・紫・茶・橙の 8 色程度であることから，今後クラスタ数を減らすことにより，より人の感覚に合ったクラスタ分けをすることができると考えられる．

次に，本研究で用いた注目物体検出法の精度について述べる．本研究で用いた画像データセットについて，注目物体を検出できたといえるものはそのうちの約 30%であった．このことから，注目物体検出手法の改良による精度の向上の余地があるといえる．

4.1 画像検索実験

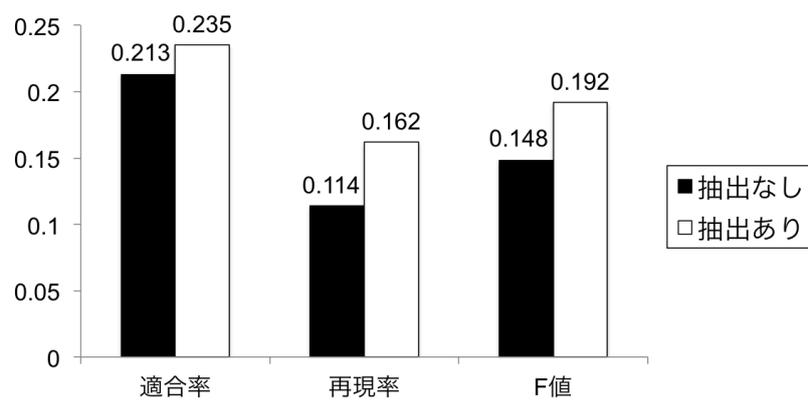


図 4.4 画像検索実験における平均適合率・平均再現率・F 値

4.1 画像検索実験

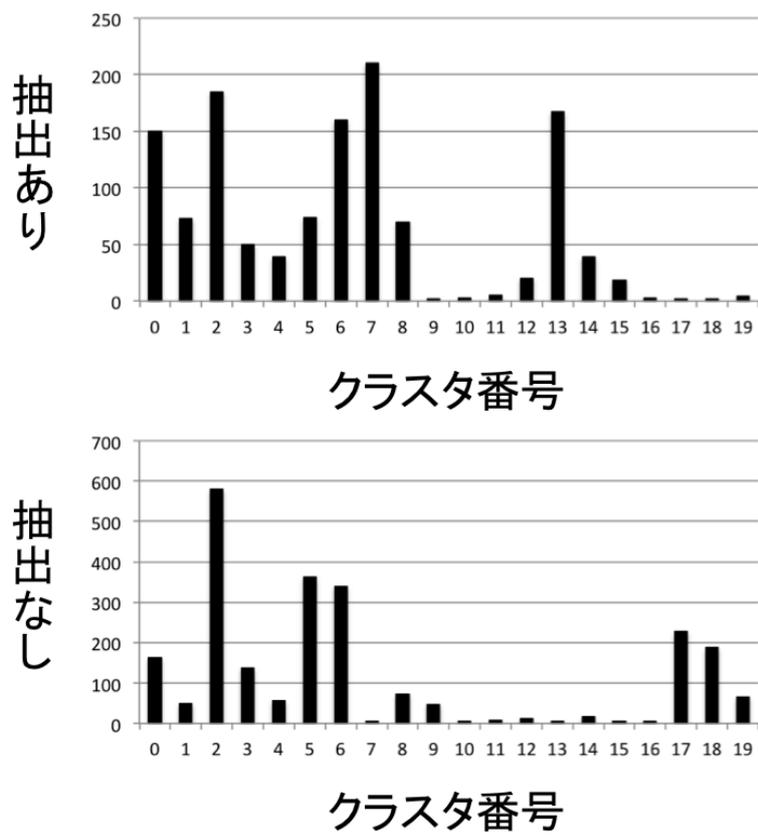


図 4.5 それぞれの手法において各クラスタに属する部分画像の枚数

第 5 章

結論

本研究では、部分領域に基づく類似画像検索において、注目物体抽出を前処理として適用し、抽出を行わない場合とで、検索精度の比較を行った。その結果、平均再現率において、抽出なしと比較して抽出ありの方が 5 ポイント高かった。このことから、類似画像検索において注目物体を抽出して特徴量を取ることが有効であると考えられる。特に、実験で用いたデータセットのように画像中に一つだけ物体がある場合や、特定の物体に対して背景にさまざまな色がありえる場合には、抽出ありとなしを比較したときに、得られる部分画像の色ヒストグラムに違いが出ることから、注目物体抽出によってより良い検索結果を得られると考えられる。しかし一方で、本研究で用いた注目物体抽出の手法は、画像中の色コントラストに基づいたものであるため、背景と注目物体の色が似通っている場合には、注目物体を抽出するのが困難となると考えられる。また、注目物体が画像中に占める割合が大きい場合にも、抽出ありとなしのときとで得られる部分画像に違いが生じにくいいため、注目物体を抽出する意義が低いと予想される。

今後は、特定のカテゴリの物体を含む画像を検索したい場合にどのような特徴量を用いるのが効果的であるかということや、別の特徴量について人間の識別しやすいクラス数について検討したい。

謝辞

本論文は著者が高知工科大学 情報学群 情報と人間専攻に在籍中の研究成果をまとめたものである。同学群准教授である吉田真一先生には指導教員として、本研究の遂行にあたり終始ご指導をいただいた。また、同学群教授の篠森敬三先生並びに同学群准教授の栗原徹先生には、副査として様々な点においてご助言をいただいた。ここに深謝の意を表する。

同研究室の笹谷氏には、システムの実装から実験、発表にいたるまで様々な場面でお力添えいただいた。また、辛い時に励まし助けていただいた。首藤氏は、常に何事にも真面目に取り組み結果を出しており、その姿勢に感銘を受けた。鈴木氏は、同じ画像検索というテーマに取り組み、システムの実装面などにおいて助けていただいた。中山氏は、プレゼンで分かりやすくご自分の研究を説明されており、その分かりやすさを自身の発表の目標とさせていただいた。佐々木氏とは苦しいときに励まし合い、また、瀬戸際まで粘り強く作業し目的を達成する姿に力づけられた。二・三年生は、和気あいあいと、時には黙々と課題や作業に取り組んでおり、研究室に活気を与えてくれた。ここに感謝の意を表する。

最後に、大学で学ぶ機会を与えてくれた家族に感謝する。

参考文献

- [1] 総務省 情報通信政策研究所 調査研究部編, “インターネット検索エンジンの現状と市場規模等に関する調査,”
<http://www.soumu.go.jp/iicp/chousakenkyu/data/research/survey/telecom/2009/2009-I-09.pdf>, (参照 2017-02-09).
- [2] Chang, Ning-San, and King-Sun Fu. “Query-by-pictorial-example,” *IEEE Transactions on Software Engineering*, vol. 6, pp. 519–524, 1980.
- [3] Karpathy, Andrej, and Li Fei-Fei. “Deep visual-semantic alignments for generating image descriptions,” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137. 2015.
- [4] Hirata, Kyoji, and Toshikazu Kato. “Query by visual example,” *In International Conference on Extending Database Technology*, pp. 56–71. Springer Berlin Heidelberg, 1992.
- [5] Rui, Yong, Thomas S. Huang, and Shih-Fu Chang. “Image retrieval: Current techniques, promising directions, and open issues,” *Journal of visual communication and image representation*, vol. 10, no. 1, pp. 39–62, 1999.
- [6] M. Serata, Y. Hatakeyama, and K. Hirota, “Designing Image Retrieval System with the Concept of Visual Keys,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 10, no. 2, pp. 136–144, 2006.
- [7] Ward Jr, Joe H. “Hierarchical grouping to optimize an objective function,” *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.
- [8] T. Ojala, M. Pietikinen, and D. Harwood, “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,” *Proceedings of the 12th IAPR International Conference on Pattern Recognition*

参考文献

- (*ICPR 1994*), vol. 1, pp. 582–585, 1994.
- [9] Liu, Guang-Hai, Zuo-Yong Li, Lei Zhang, and Yong Xu. “Image retrieval based on micro-structure descriptor,” *Pattern Recognition*, vol. 44, no. 9, pp. 2123–2133, 2011.
- [10] Feng, Lin, Jun Wu, Shenglan Liu, and Hongwei Zhang. “Global correlation descriptor: a novel image representation for image retrieval,” *Journal of Visual Communication and Image Representation*, vol. 33, pp. 104–114, 2015.
- [11] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu, “Global Contrast Based Salient Region Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.
- [12] Koch, Christof, and Shimon Ullman. “Shifts in selective visual attention: towards the underlying neural circuitry,” *In Matters of intelligence*, pp. 115–141. Springer Netherlands, 1987.
- [13] Robinson, D.L. and Petersen, S.E., “The pulvinar and visual salience,” *Trends in neurosciences*, vol. 15, no. 4, pp.127–132, 1992.
- [14] L. Itti, C. Koch and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [15] Shen, Xiaohui, and Ying Wu. “A unified approach to salient object detection via low rank matrix recovery,” *In Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, pp. 853–860. IEEE, 2012.
- [16] Y. Shiozaki, and S. Yoshida, “An Attentional Region Detection Method for Extracting Visual Keys in Visual Key Image Retrieval,” *International Conference on Next Era Information Networking*, pp. 231–234, 2008.
- [17] Rother, Carsten, Vladimir Kolmogorov, and Andrew Blake. “Grabcut: Interactive foreground extraction using iterated graph cuts,” *In ACM transactions on graphics*

参考文献

- (*TOG*), vol. 23, no. 3, pp. 309–314. ACM, 2004.
- [18] P. Felzenszwalb and D. Huttenlocher, “Efficient graph-based image segmentation,” *International journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [19] Image Segmentation, <http://cs.brown.edu/~pff/segment/>, (参照 2017-02-08).
- [20] GitHub - MingMingCheng/CmCode: Ming-Ming Cheng’s open source projects, 2016, <https://github.com/MingMingCheng/CmCode>, (参照 2017-02-08).
- [21] “ImageNet”, <http://www.image-net.org/>, (参照 2017-02-10).